

# Analytical continuation of imaginary axis data using maximum entropy

O. Gunnarsson<sup>(1)</sup>, M. W. Haverkort<sup>(1)</sup> and G. Sangiovanni<sup>(2)</sup>

<sup>1</sup>*Max-Planck-Institut für Festkörperforschung, D-70506 Stuttgart, Germany*

<sup>2</sup>*Institut für Festkörperphysik, Technische Universität Wien, Vienna, Austria*

We study the maximum entropy (MaxEnt) approach for analytical continuation of spectral data from imaginary times to real frequencies. The total error is divided in a statistical error, due to the noise in the input data, and a systematic error, due to deviations of the default function, used in the MaxEnt approach, from the exact spectrum. We find that the MaxEnt approach in its classical formulation can lead to a nonoptimal balance between the two types of errors, leading to an unnecessary large statistical error. The statistical error can be reduced by splitting up the data in several batches, performing a MaxEnt calculation for each batch and averaging. This can outweigh an increase in the systematic error resulting from this approach. The output from the MaxEnt result can be used as a default function for a new MaxEnt calculation. Such iterations often lead to worse results due to an increase in the statistical error. By splitting up the data in batches, the statistical error is reduced and the increase resulting from iterations can be outweighed by a decrease in the systematic error. Finally we consider a linearized version to obtain a better understanding of the method.

## I. INTRODUCTION

The analytical continuation of spectral functions from imaginary time  $\tau$  to real energies  $\omega$  is a difficult problem due to its ill-posed nature, i.e., the output can depend very sensitively on the input. For strongly correlated electrons, however, this is an important problem. Most approaches for such systems involve uncontrolled approximations. Using quantum Monte-Carlo (QMC) methods or quantum cluster methods it is, however, possible to obtain accurate data for Green's functions and response functions on the imaginary axis, raising the problem of analytical continuation to the real axis. Since these methods provide data with statistical noise, the ill-posed nature of the problem makes analytical continuation very difficult.

This problem can be treated within the Bayesian theory.<sup>1,2</sup> The problem is regularized by introducing an entropy in terms of the deviation of the output real axis spectrum from some default function. The importance of the entropy is controlled by a parameter  $\alpha$ , which is determined using statistical arguments.<sup>1,2</sup> This method is referred to as the Maximum Entropy (MaxEnt) method. It has been rather successful in performing analytical continuations. Alternative methods have been proposed, such as Padé approximations,<sup>3,4</sup> singular value decomposition,<sup>5</sup> stochastic regularization<sup>6</sup> and sampling schemes.<sup>7,8</sup>

In this paper we focus on the MaxEnt method. This method is usually discussed in terms of the Bayesian theory. Here we start from the equations generated by the MaxEnt formalism and use an algebraic approach to analyze the theory. We discuss the accuracy that can be obtained within this framework. The error in the output spectral function can be split up in a statistical error, due to the noise in the input data, and a systematic error, due to the deviation of the default function from the true spectrum. The choice of  $\alpha$  determines the relative size of these errors. In the classical MaxEnt method the

most probable  $\alpha$  is chosen.<sup>1</sup> We find that this choice can make the statistical error unnecessary large.

The input data is typically given as a number  $N_{\text{sample}}$  of samples,  $\bar{G}_\nu(\tau)$ , where each sample gives a (noisy) version of the imaginary time function  $G(\tau)$ . We find that the accuracy can sometimes be improved by splitting up the samples in  $N_{\text{calc}}$  sub sets (batches), with  $N_{\text{sample}}/N_{\text{calc}}$  samples in each batch. We then perform  $N_{\text{calc}}$  MaxEnt calculations, each with  $N_{\text{sample}}/N_{\text{calc}}$  samples, and then average the results, instead of performing one MaxEnt calculation  $N_{\text{sample}}$  samples. This approach reduces the statistical error at the cost of an increase in the systematic error.

We also discuss the possibility of an iterative MaxEnt method, where the output is used to define a new default function. This usually works poorly, and we show that this is due to an increase in the statistical error, overwhelming the improvement in the systematic error. However, if the data are split in batches, as discussed above, the importance of the statistical error can be reduced to the point where the approach improves the total accuracy.

To further analyze the results, we introduce an alternative method with a new, slightly different definition of the entropy. This leads to a set of linear equations, where the propagation of the errors can be analyzed more easily and features of the MaxEnt method better understood. This method, however, does not guarantee a positive spectral function, and it is less useful for practical calculations.

In this paper we focus on a response function, the optical conductivity  $\sigma(\omega)$ . We introduce a typical  $\sigma(\omega)$ , which in the following will be referred to as the “exact”  $\sigma(\omega)$ . The form of  $\sigma(\omega)$  was chosen using results for the two-dimensional Hubbard model as a guide. This model of  $\sigma(\omega)$  can easily and accurately be transformed to imaginary axis data. We add statistical noise to the data and then transform the data back to the real axis, using the various modifications of the MaxEnt method. If a given method worked perfectly the  $\sigma(\omega)$  that we started with

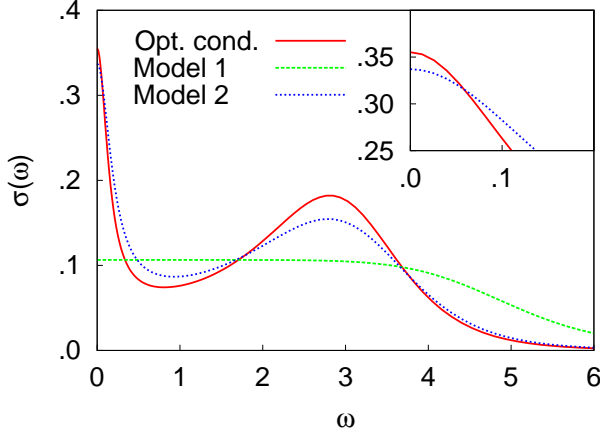


FIG. 1: (color on-line) “Exact” spectral function and two different default models as a function of frequency. The inset shows the models on a small energy scale.

should be recovered exactly. The deviations from the “exact”  $\sigma(\omega)$  are then a measure of the accuracy of the different approaches.

As an example, Fig. 1 shows an “exact” optical conductivity and two default models used in the MaxEnt approach. The optical conductivity has a Drude like peak at  $\omega = 0$  and a “Hubbard” peak at  $\omega \sim 3$  corresponding to transitions between the Hubbard bands. The default models are chosen so that they satisfy the exact sum rule. The two models are chosen according to two different strategies. It is sometimes argued that the default model should contain little information, apart from certain exact results, such as sum rules. This way the results are not prejudiced by possible incorrect assumptions. Model 1 has been chosen this way. Alternatively, as a default model one can use the output from a calculation at a higher temperature  $T$ . Model 2 has therefore been chosen to be quite similar to the “exact” result, but with all features somewhat broader. Model 2 will naturally deliver much more accurate output spectra.

The paper is organized as follows. In Sec. II we introduce the formalism. Sec. III describes how a MaxEnt calculation is performed as an average of several MaxEnt calculations and Sec. IV discusses an iterative MaxEnt method. In Sec. V we present a simplified entropy definition, leading to linear equations.

## II. FORMALISM

We introduce the basic formalism, essentially following Jarrell and Gubernatis,<sup>1</sup> and then provide error estimates. The function  $G_i = G(\tau_i)$  for imaginary time  $\tau_i$  is related to a spectral function  $A_i = A(\omega_i)$  on the real frequency axis  $\omega$ ,

$$G_i = \sum_{j=1}^{N_\omega} K_{ij} A_j \quad i = 1, \dots, N_\tau, \quad (1)$$

via a kernel  $K_{ij} = K(\tau_i, \omega_j)$ , given for some discrete values  $\omega_j$  of  $\omega$ . For the case of the optical conductivity, considered here, the kernel is given by

$$K_{ij} = \frac{1}{\pi} \frac{\omega_j}{1 - \exp(-\beta\omega_j)} (e^{-\omega_j\tau_i} + e^{-(\beta-\tau_i)\omega_j}) f_j, \quad (2)$$

where  $f_j$  is a weight factor chosen so that Eq. (1) corresponds to an integral over  $\omega$ . For the electron Green’s function the corresponding kernel is

$$K_{ij} = \frac{e^{-\tau_i\omega_j}}{1 + e^{-\beta\omega_j}} f_j. \quad (3)$$

We introduce a likelihood function

$$L = \frac{1}{2} \sum_{i=1}^{N_\tau} \left( \frac{\bar{G}_i - G_i}{\sigma_i} \right)^2, \quad (4)$$

where  $\bar{G}_i$  are data obtained from, e.g., a Monte-Carlo calculation, with the statistical accuracy  $\sigma_i$ , and  $G_i$  has been calculated from Eq (1). We also introduce the entropy

$$S = \sum_{i=1}^{N_\omega} f_i (A_i - m_i - A_i \ln \frac{A_i}{m_i}), \quad (5)$$

where  $m_i$  is a default model. The quantity  $L - \alpha S$  is then minimized with respect  $A_j$ . This leads to the equations

$$- \sum_{i=1}^{N_\tau} \frac{\bar{G}_i - G_i}{\sigma_i^2} K_{ij} + \alpha f_j \ln \frac{A_j}{m_j} = 0 \quad (6)$$

These equations are solved to obtain the spectral function  $A_i$ . The quantity  $\alpha$  can be determined using statistical methods, giving the most probable  $\alpha$ . This is referred to as the classical MaxEnt method.<sup>1</sup> Alternatively, one can average the spectrum calculated for different values of  $\alpha$ , using the probability of that  $\alpha$  as a weighting function.<sup>1</sup> This method, Bryan’s method, gives similar results for the cases considered here.

To estimate the error in this approach, we express the calculated spectral function  $A$  in terms of the exact result  $A^{\text{exact}}$  as

$$A_i = A_i^{\text{exact}} + \Delta A_i, \quad (7)$$

where  $\Delta A_i$  is the error in  $A_i$ . We assume that the error is sufficiently small that the logarithm in Eq. (6) can be expanded to lowest order. Then

$$- \sum_{i=1}^{N_\tau} \frac{\Delta \bar{G}_i - \Delta G_i}{\sigma_i^2} K_{ij} + \alpha f_j \left( \ln \frac{A_j^{\text{exact}}}{m_j} + \frac{\Delta A_j}{A_j^{\text{exact}}} \right) = 0, \quad (8)$$

where  $\Delta G_i = \sum_j K_{ij} \Delta A_j$  and  $\Delta \bar{G}_i = \bar{G}_i - \sum_j K_{ij} A_j^{\text{exact}}$  is the error in  $\bar{G}_i$  due to the statistical noise. To solve these equations, we define

$$a_j = \sum_{i=1}^{N_\tau} \frac{\Delta \bar{G}_i K_{ij}}{\sigma_i^2} + \alpha f_j \ln \frac{m_j}{A_j^{\text{exact}}} \quad (9)$$

$$b_{jk} = \sum_{i=1}^{N_\tau} \frac{K_{ij}K_{ik}}{\sigma_i^2} + \frac{\alpha f_j}{A_j^{\text{exact}}} \delta_{jk}. \quad (10)$$

Using matrix notations,

$$\Delta A = b^{-1}a = b^{-1}K^T \sigma^{-2} \Delta \bar{G} + b^{-1} \alpha f \ln\left(\frac{m}{A^{\text{exact}}}\right). \quad (11)$$

The error  $w$  is defined as

$$w = \left\langle \sum_{i=1}^{N_\omega} (\Delta A_i)^2 f_i \right\rangle \equiv w_{\text{stat}} + w_{\text{syst}}, \quad (12)$$

where  $\langle \dots \rangle$  denotes the average over many different realizations of the noise  $\Delta \bar{G}_i$  in the input data. Here  $w_{\text{stat}}$  is the error due to this noise and  $w_{\text{syst}}$  is the error due to the deviation

$$\Delta m_i = m_i - A_i^{\text{exact}} \quad (13)$$

of the default function from the exact result. Since the noise is random, there is no contribution from the cross term in the square of the two terms in Eq. (11). The statistical error can then be written as

$$\begin{aligned} w_{\text{stat}} &= \left\langle \sum_{j,k=1}^{N_\tau} \frac{\Delta G_j}{\sigma_j^2} (K b^{-1} f b^{-1} K^T)_{jk} \frac{\Delta G_k}{\sigma_k^2} \right\rangle \\ &= \sum_{j=1}^{N_\tau} \frac{(K b^{-1} f b^{-1} K^T)_{jj}}{\sigma_j^2} = \text{Tr} \sigma^{-2} K b^{-1} f b^{-1} K^T. \end{aligned} \quad (14)$$

The second equality was obtained by noticing that the terms  $j \neq k$  do not contribute on the average and that the average of  $\Delta G_i^2$  is  $\sigma_i^2$ . Later we consider the average over  $N_{\text{calc}}$  MaxEnt calculations, each using data with the statistical accuracy  $\sigma$ . The statistical error  $w_{\text{stat}}$  is then reduced by a factor of  $N_{\text{calc}}$ , since  $w_{\text{stat}}$  refers to the square of the error in the output spectrum.

For the systematic error we obtain

$$w_{\text{syst}} = \ln\left(\frac{m}{A^{\text{exact}}}\right) \alpha b^{-1} f b^{-1} \alpha f \ln\left(\frac{m}{A^{\text{exact}}}\right). \quad (15)$$

The results in Eqs. (14, 15) apply to the case when MaxEnt calculation is not iterated. For an iterative calculation we use Eq. (22) below.

Fig. 2 shows  $w_{\text{stat}}$  and  $w_{\text{syst}}$  as a function of  $\alpha$  for different  $\sigma$ . The figure illustrates that  $w_{\text{stat}}$  behaves approximately as  $1/\alpha$ . This illustrates the importance of introducing entropy, i.e., using an  $\alpha > 0$ . For  $\alpha = 0$ , the matrix  $b^{-1}$  is ill-behaved and the statistical error would be huge. Since  $w_{\text{stat}}$  depends only weakly on  $\sigma$ , it is not possible to make  $w_{\text{stat}}$  small for  $\alpha = 0$  by simply reducing  $\sigma$  (within reasonable limits). Introducing  $\alpha > 0$  regularizes  $b$  and leads to a manageable statistical error. The systematic error increases with  $\alpha$  and there is therefore an optimal value of  $\alpha$  where the total error is minimum. The dependence of the systematic error on  $\sigma$  is shown in Fig. 3. It behaves roughly as  $\sqrt{\sigma}$ . The optimal  $\alpha$  therefore increases as  $\sigma$  is reduced. It also increases as the

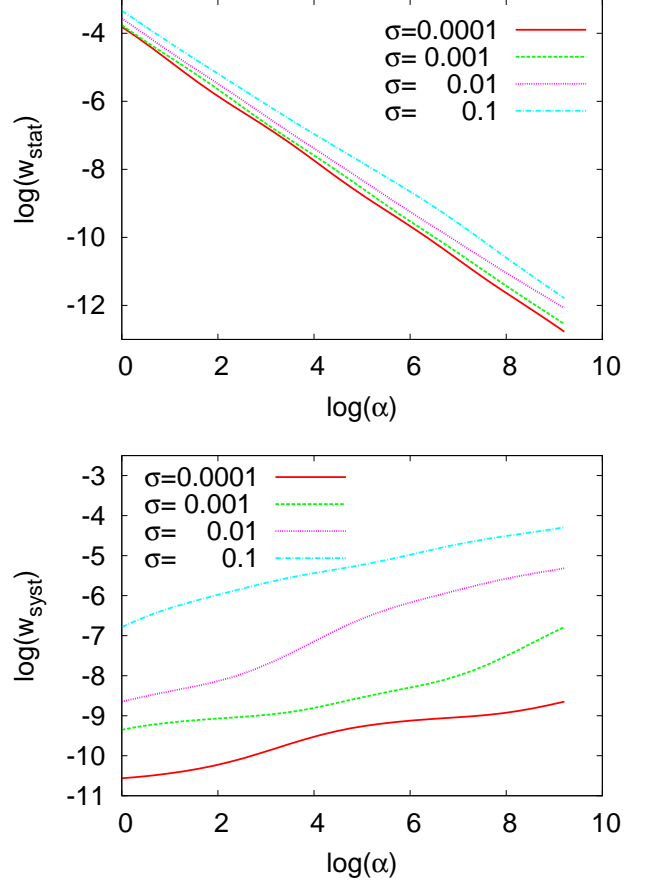


FIG. 2: (color on-line) Statistical ( $w_{\text{stat}}$ ) [Eq. (14)] and systematic ( $w_{\text{syst}}$ ) [Eq. (15)] errors for default model 1 as a function of  $\alpha$  and for different values of  $\sigma$ . The parameters are  $\beta = 15$  and  $N_\tau = 60$ .

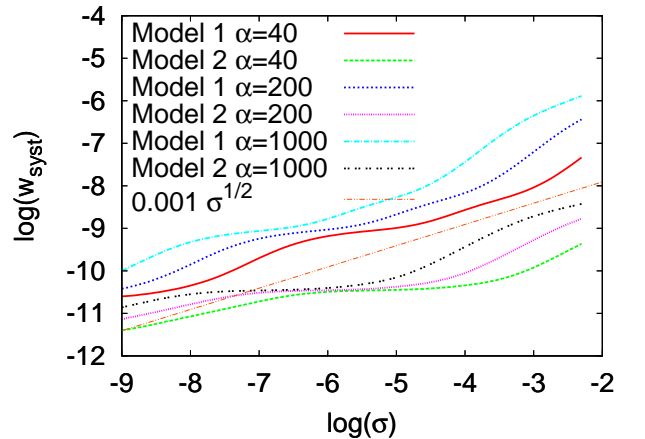


FIG. 3: (color on-line) Systematic error ( $w_{\text{syst}}$ ) [Eq. (15)] for default models 1 and 2 as a function of  $\sigma$  and for different values of the  $\alpha$ . The straight line shows the curve  $0.001\sigma^{1/2}$ , illustrating that  $w_{\text{stat}}$  is approximately proportional to  $\sqrt{\sigma}$ . The parameters are  $\beta = 15$  and  $N_\tau = 60$ .

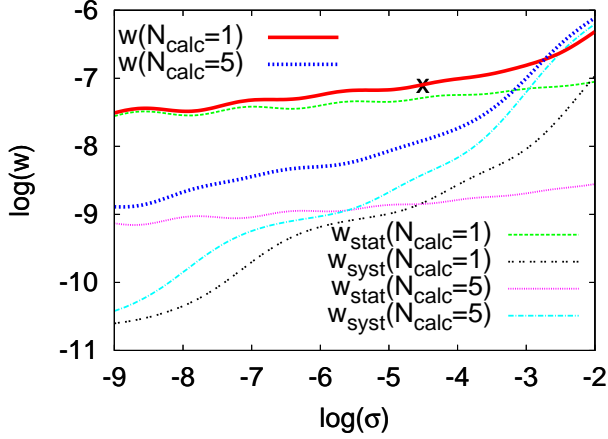


FIG. 4: (color on-line) Statistical ( $w_{\text{stat}}$ ), systematic ( $w_{\text{syst}}$ ) and total ( $w$ ) errors in a MaxEnt calculation for the spectrum in Fig. 1, default model 1,  $\alpha = 40$  and  $\beta = 15$  and 100 samples, each with the accuracy  $\sigma$ . The full thick (red) line shows  $w$  when one ( $N_{\text{calc}} = 1$ ) MaxEnt calculation is performed for the average of over all data and the thick broken (blue) line the result when  $N_{\text{calc}} = 5$  MaxEnt calculations are averaged, each calculation using the average of  $100/N_{\text{calc}}$  samples. The cross corresponds to a historic MaxEnt calculation for  $\sigma = 0.01$ , which gives  $\alpha \approx 40$ , used in the figure. The thick broken (blue) curve illustrates that a substantially lower error can be obtained by averaging 5 MaxEnt calculations.

default model is made more accurate, e.g., by replacing default model 1 by model 2.

Since  $\omega = 0$  is often of particular interest we use a logarithmic  $\omega$ -mesh. For the case  $0 \leq \omega \leq \omega_{\text{max}}$  we use

$$\omega_i = \exp[(i-1)dx + \ln\gamma] - \gamma, \quad (16)$$

where  $dx = [\ln(\omega_{\text{max}} + \gamma) - \ln\gamma]/(N_\omega - 1)$ . A small value of  $\gamma$  leads to a smaller spacing of the points close to  $\omega = 0$ . We have typically used  $N_\omega = 121$  points,  $\gamma = 0.5$ ,  $\omega_{\text{max}} = 12$ ,  $\beta = 15$  and  $N_\tau = 60$ . For simplicity, we assume that the statistical error is given by

$$\sigma_i = G_i\sigma, \quad (17)$$

in terms of some overall accuracy  $\sigma$ . To perform these calculations we have developed a MaxEnt code, which was found to give almost identical results to a code made available to us by Jarrell.<sup>1</sup>

### III. MULTIPLE MAXENT CALCULATIONS

A QMC calculation is arranged so that it gives a number of samples,  $\bar{G}_\nu(\tau_i)$ , of  $G(\tau_i)$ . From these data one can calculate the statistical accuracy  $\sigma_i$ , check if the data are Gaussian and check for (undesirable) correlations between the noise for different values of  $\tau$ .<sup>1</sup> The data  $\bar{G}_\nu(\tau_i)$  are then averaged over  $\nu$  to obtain  $\bar{G}(\tau_i)$  that is the input for the MaxEnt calculation, possibly after removing correlations between the noise at different  $\tau$ -points.<sup>1</sup>

TABLE I: Statistical ( $w_{\text{stat}}$ ), systematic ( $w_{\text{syst}}$ ) and total ( $w$ ) error in MaxEnt calculations for the spectrum in Fig. 1 and default models 1 or 2. 100 samples, each with the accuracy  $\sigma$ , were split up in  $N_{\text{calc}}$  batches with  $100/N_{\text{calc}}$  samples and used in  $N_{\text{calc}}$  calculations. The average of the output was used as a default model, performing  $N_{\text{iter}}$  iterations. The errors were obtained from Eqs. (14, 15) for  $N_{\text{iter}} = 1$  and from Eq. (22) for  $N_{\text{iter}} > 1$ . The parameters were  $\beta = 15$  and  $N_\tau = 60$ .

Default model	$\alpha$	$N_{\text{calc}}$	$N_{\text{iter}}$	$w_{\text{stat}}$	$w_{\text{syst}}$	$w$
1	40	1	1	$6.7 \cdot 10^{-4}$	$1.4 \cdot 10^{-4}$	$8.1 \cdot 10^{-4}$
1	40	5	1	$1.4 \cdot 10^{-4}$	$2.1 \cdot 10^{-4}$	$3.5 \cdot 10^{-4}$
1	40	1	2	$14 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$15 \cdot 10^{-4}$
1	40	100	4	$0.4 \cdot 10^{-4}$	$0.9 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$
2	720	1	1	$4.2 \cdot 10^{-5}$	$4.4 \cdot 10^{-5}$	$8.6 \cdot 10^{-5}$
2	720	2	1	$2.3 \cdot 10^{-5}$	$5.6 \cdot 10^{-5}$	$7.9 \cdot 10^{-5}$
2	720	1	2	$9.7 \cdot 10^{-5}$	$3.3 \cdot 10^{-5}$	$13 \cdot 10^{-5}$
2	720	20	8	$2.4 \cdot 10^{-5}$	$3.9 \cdot 10^{-5}$	$6.3 \cdot 10^{-5}$

We now consider the case where we have 100 samples, each with the accuracy  $\sigma$ . After averaging over all the samples, the accuracy of the resulting data is  $\sigma/\sqrt{100} = \sigma/10$ . The value of  $\alpha$  in a classical MaxEnt calculation depends on the specific realization of the noise. We therefore perform many calculations, each with a different realization of the noise, and average over  $\alpha$ . For  $\sigma = 0.01$ ,  $\beta = 15$  and  $N_\tau = 60$ , classical MaxEnt calculations using the spectrum in Fig. 1 and the default model 1 then gave on the average  $\alpha \approx 40$ .

Fig. 4 shows the statistical and systematic errors for a fixed  $\alpha = 40$  as a function of  $\sigma$  ( $N_{\text{calc}} = 1$ ). The cross gives the total error of a classical MaxEnt calculation corresponding to 100 samples with  $\sigma = 0.01$ . The figure illustrates that the statistical error is much larger than the systematic error for the MaxEnt calculation. This is also illustrated in Fig. 5a. This shows the result of 20 MaxEnt calculations with different realizations of the noise, each with 100 samples with the accuracy  $\sigma$ . The thick (red) line shows the exact spectrum. The calculated spectra (thin green lines) scatter strongly around the exact result, illustrating a large statistical error. On the average, these spectra also deviate somewhat from the exact result, the value of  $\sigma(0)$  being slightly too small and the Hubbard peak being somewhat shifted towards lower energies, illustrating a small systematic error.

We next group the 100 samples in  $N_{\text{calc}} = 5$  batches, each with 20 samples, and perform  $N_{\text{calc}}$  MaxEnt calculations. The accuracy of the data in these MaxEnt calculations is then only  $\sqrt{N_{\text{calc}}}\sigma/10$ . This increases both the systematic and statistical errors somewhat. Averaging these calculation, however, reduces the statistical error by a factor  $N_{\text{calc}}$ . In Fig. 4 this leads to a large net reduction in the statistical error, which more than compensates for the increase of the systematic error. This is illustrated in Fig. 5b, which shows 20 such results, each one obtained by averaging  $N_{\text{calc}} = 5$  MaxEnt calculations with  $100/N_{\text{calc}}$  samples, but with different re-

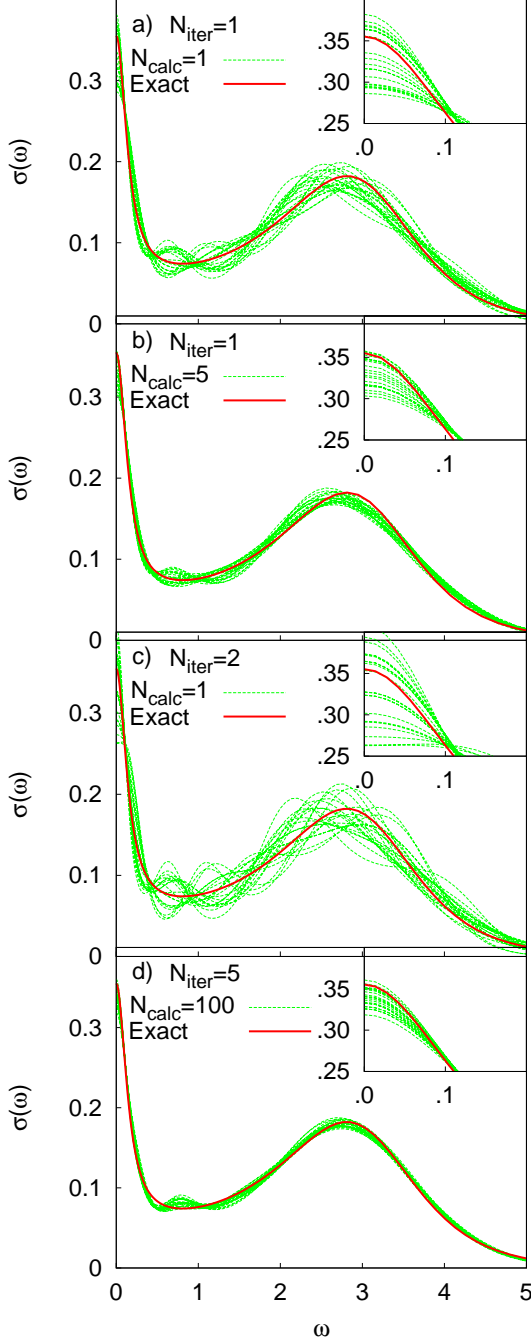


FIG. 5: (color on-line) Optical conductivity calculated for the default model 1 using different methods. 100 samples, each with the accuracy  $\sigma = 0.01$  were given. a) Each curve shows results of a classical MaxEnt calculation using an average of all 100 samples. The figure shows 20 such curves, each corresponding to a different realization of the noise. b) Each curve shows the average of  $N_{\text{calc}} = 5$  MaxEnt calculations using  $100/N_{\text{calc}}$  samples. c) Each curve shows the results of iterating the calculations in a) once, using the output in a) as a default function in the next MaxEnt calculation. d) Each curve shows the results of iterating MaxEnt calculations  $N_{\text{iter}} = 5$  times.  $N_{\text{calc}} = 100$  was used, and the default function was obtained from the average of these  $N_{\text{calc}}$  calculations. The parameters were  $\beta = 15$ ,  $N_{\tau} = 60$  and  $\alpha = 40$ .

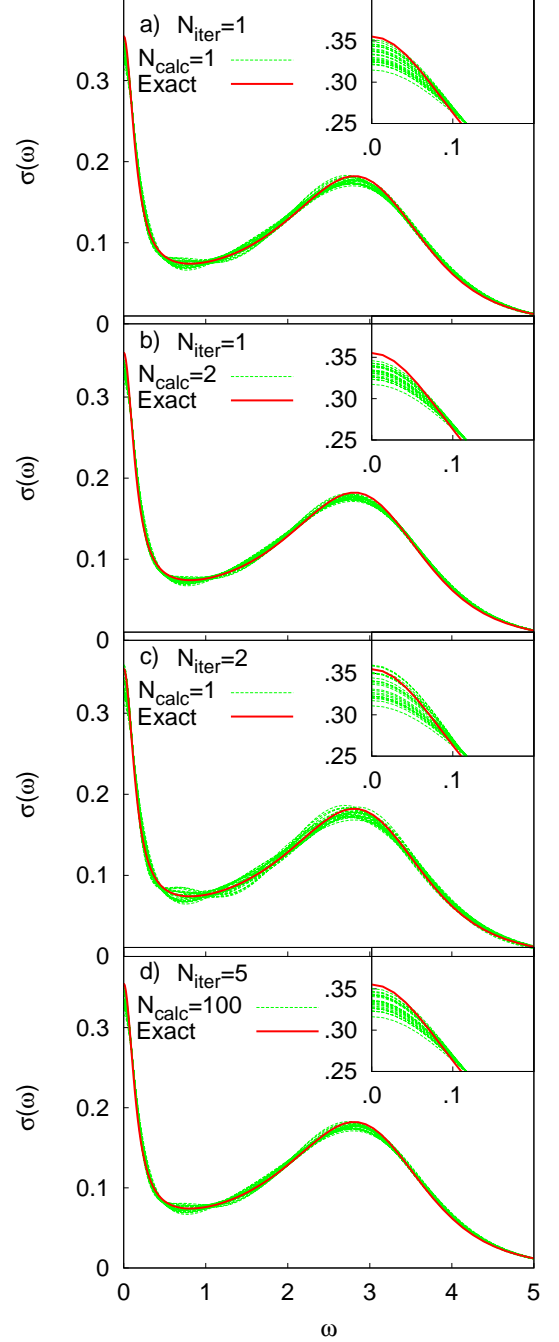


FIG. 6: (color on-line) The same as Fig. 5, but starting from the default model 2 and using  $\alpha = 720$ .

alizations of the noise. The spread between the curves is substantially smaller ( $w_{\text{stat}} = 0.00014$  vs.  $0.00067$ ) than in Fig. 5a, while the systematic error is somewhat larger ( $w_{\text{syst}} = 0.00021$  vs.  $0.00014$ ). This leads to a substantial improvement in the total error ( $w = 0.00035$  vs.  $0.00081$ ). These results are also shown in Table I.

The reason for this improvement is that that  $w_{\text{stat}} \gg w_{\text{syst}}$  in the MaxEnt calculation with  $N_{\text{calc}} = 1$  and that  $w_{\text{stat}}$  and  $w_{\text{syst}}$  have different dependencies on  $N_{\text{calc}}$ . For

$\alpha = 0$ , Eq. (14) gives that  $w_{\text{stat}} \sim \sigma^2$ . Splitting up the calculation in  $N_{\text{calc}}$  calculations makes the effective  $\sigma$  a factor  $\sqrt{N_{\text{calc}}}$  larger, while averaging reduces the error by a factor  $N_{\text{calc}}$ . The net result would be an unchanged statistical error. It is therefore crucial that the method has been regularized by introducing an entropy. Fig. 4 shows that for realistic values of  $\alpha$ ,  $w_{\text{stat}}$  actually has a quite weak dependence on  $\sigma$ , rather than behaving as  $\sigma^2$ . Splitting up the samples in several batches, and thereby reducing the accuracy of each batch, leads to a small increase in  $w_{\text{stat}}$  for each individual calculation. The averaging over  $N_{\text{calc}}$  calculations, however, reduces  $w_{\text{stat}}$  by a factor  $N_{\text{calc}}$ . At the same time  $w_{\text{syst}}$  is increased, but only by approximately a factor  $N_{\text{calc}}^{1/4}$ , since this quantity behaves approximately as  $\sqrt{\sigma}$  and  $\sigma$  increases by a factor  $\sqrt{N_{\text{calc}}}$ .

Fig. 6 and Table I show the corresponding results using default model 2. In this case the classical MaxEnt calculation chooses a value of  $\alpha$  that makes  $w_{\text{stat}}$  and  $w_{\text{syst}}$  comparable. The gain from splitting up the samples and performing several MaxEnt calculations is then much smaller.

We are now in the position to discuss the limits of accuracy that can be obtained in this approach. We consider as before 100 samples with the accuracy  $\sigma = 0.01$  and allow for any combination of  $\alpha$ ,  $N_{\text{calc}} \leq 100$  and  $N_{\text{iter}} \leq 40$ . Starting from default model 1, we obtain the results shown in Fig. 7a. The curve “One calc.” shows the result of a traditional MaxEnt calculation, using all the samples in one calculation ( $N_{\text{calc}} = 1$  and  $N_{\text{iter}} = 1$ ). If a classical MaxEnt calculation is performed,  $\alpha \approx 20$  is obtained. This result is shown by a cross. We can see that this value of  $\alpha$  is not optimal, and a larger  $\alpha$  would have given a smaller error. We next allow for  $N_{\text{calc}} > 1$  calculations, each using  $100/N_{\text{calc}}$  samples. We find the value of  $N_{\text{calc}}$  which gives the best agreement with the “exact”  $\sigma(\omega)$ . This (“Several opt.”) leads to a much higher accuracy for small values of  $\alpha$ . The curve is almost flat as a function of  $\alpha$  over a substantial range. For large values of  $\alpha$ ,  $N_{\text{calc}} = 1$  gives the best accuracy, and the curve falls on top of the curve “One calc.”.

To provide a criterion for how to split up the data in batches, we consider the statistical error [(Eq. 14)] again. As before we consider the case of  $N_{\text{sample}}$  samples, each with the accuracy  $\sigma$ , divided in  $N_{\text{calc}}$  batches with  $N_{\text{sample}}/N_{\text{calc}}$  samples in each. We define the product

$$M(\sigma) = b^{-1} K^T \sigma^{-2}, \quad (18)$$

where  $b$  also depends on  $\sigma$ . The statistical error of the  $k$ th calculation is then written as

$$\Delta A_i^{(k)} = \sum_{j=1}^{N_\tau} M_{ij}(\sigma_{N_{\text{calc}}}) \frac{N_{\text{calc}}}{N} \sum_{\nu=1}^{N/N_{\text{calc}}} \Delta \bar{G}_j^{\nu+(k-1)N/N_{\text{calc}}}, \quad (19)$$

where  $\Delta \bar{G}^\nu$  is the error in the  $\nu$ th sample, and the statistical accuracy  $\sigma_{N_{\text{calc}}} = \sigma/\sqrt{N/N_{\text{calc}}}$  enters due to the averaging over  $N/N_{\text{calc}}$  samples. We then average over

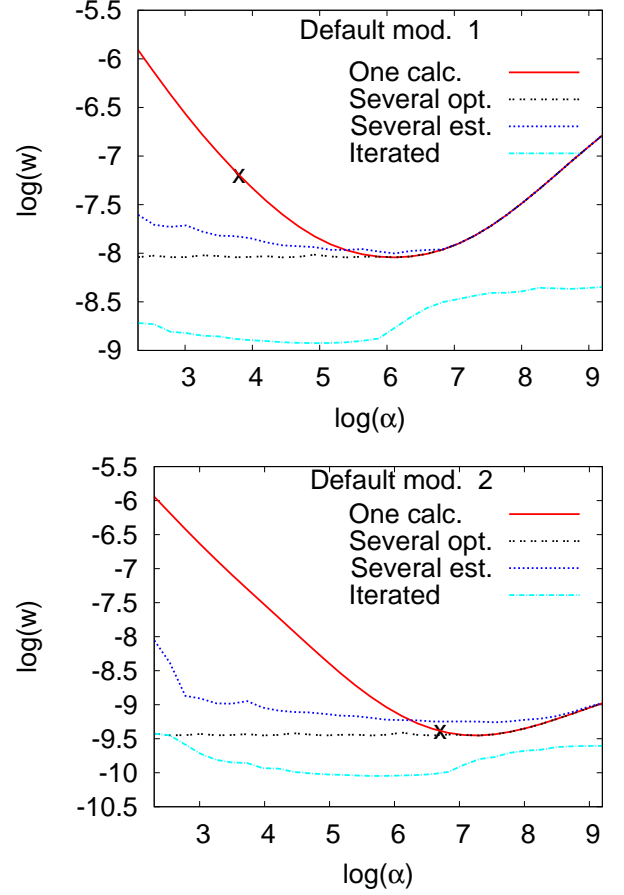


FIG. 7: (color on-line) Accuracy  $w$  of MaxEnt calculations for 100 samples, each with accuracy  $\sigma$ . “One calc.” uses the average of all samples in one MaxEnt calculation. “Several opt.” and “Several est.” split up the samples in several batches and averages the resulting MaxEnt calculations. “Several opt.” does this in the optimal way and “Several est.” uses a prescription for finding the splitting when the exact result is not known. “Iterated” in addition uses the output spectral function as default model in an iterative approach. The cross shows the result of a classical MaxEnt calculation. a) shows results for default model 1 and b) for default model 2. The parameters are  $\beta = 15$  and  $N_\tau = 60$ .

the  $N_{\text{calc}}$  calculations and obtain the error

$$\Delta A_i = \sum_{j=1}^{N_\tau} M_{ij}(\sigma_{N_{\text{calc}}}) \frac{1}{N} \sum_{\nu=1}^N \Delta \bar{G}_j^\nu. \quad (20)$$

The average difference between two calculations with  $N_{\text{calc}}$  and  $M_{\text{calc}}$  batches can then be written as

$$w_{MN} \equiv \sum_i [A_i(N_{\text{calc}}) - A_i(M_{\text{calc}})]^2 w_i = \sum_{ij} w_i \left[ \frac{M_{ij}(\sigma_{N_{\text{calc}}}) \sigma_{N_{\text{calc}}}}{\sqrt{N_{\text{calc}}}} - \frac{M_{ij}(\sigma_{M_{\text{calc}}}) \sigma_{M_{\text{calc}}}}{\sqrt{M_{\text{calc}}}} \right]^2. \quad (21)$$

This result represents an average over many different realizations of the error  $\Delta \bar{G}^{(k)}$ . In addition to the sta-



tistical contribution to the difference there is a systematic contribution due to the error in the default function. We then compare calculations with  $N_{\text{calc}} = N_{\text{sample}}$  and  $M_{\text{calc}} = N_{\text{sample}}/2$  batches. In the second calculation the statistical error is larger and the systematic error is smaller. If the total difference between the two calculations is larger than twice the expected statistical difference, this suggests that the gain in the systematic error outweighs the loss in the statistical error and the second calculation is accepted. We then compare this calculation with a calculation with  $N_{\text{sample}}/4$  batches and if the latter is favorable the procedure is continued, considering  $N_{\text{sample}}/10$ ,  $N_{\text{sample}}/20$  and  $N_{\text{sample}}/50$  batches. The resulting accuracy is shown by the curves “Several est.” in Fig. 7. This curve is above the curve “Several opt.”, but the difference is not very large for most values of  $\alpha$ .

#### IV. ITERATING MAXENT

Once a MaxEnt calculation has been performed, one can try to improve the default function by using the output spectral function as a new default function. Such an iterative approach, however, is usually not recommended. Fig. 5c shows the results of such calculations using default model 1. Indeed, the spread between different calculations is larger than in the noniterated case in Fig. 5a, implying an increased statistical error.

Eqs. (14, 15) used to calculate the statistical and systematic errors for a noniterated default model are not appropriate in the case of iterations. The reason is that the default model in this case contains statistical errors due to the iteration procedure. Instead we perform many calculations  $N$  of the type shown in Fig. 5c, giving spectral functions  $A_i^\nu$ ,  $\nu = 1, \dots, N$ . We then calculate

$$\begin{aligned} A_i^{\text{av}} &= \frac{1}{N} \sum_{\nu=1}^N A_i^\nu \\ w_{\text{stat}} &= \frac{1}{N-1} \sum_{\nu=1}^N \sum_{i=1}^{N_\omega} (A_i^\nu - A_i^{\text{av}})^2 f_i \\ w_{\text{syst}} &= \sum_{i=1}^{N_\omega} (A_i^{\text{av}} - A_i^{\text{exact}})^2 f_i \end{aligned} \quad (22)$$

Due to nonlinearity, some of the statistical error actually shows up as a systematic error in Eq. (22), but this is neglected in the following.

The results in Table I, shows that the statistical error is more than doubled after one iteration, while the systematic error is not correspondingly reduced.

We next consider the case when the samples are split up in  $N_{\text{calc}}$  batches and default model 1 is used. Using  $N_{\text{calc}} = 100$  and  $N_{\text{iter}} = 4$  the total error is reduced, as is illustrated in Fig. 5d and Table I. By using  $N_{\text{calc}} = 100$ , we drastically reduce the statistical error. The following iterations increase the statistical error by a substantial factor, but it nevertheless remains small. At the same

time the iterations reduce the systematic error, so that both are improved compared with the noniterated case. Fig. 6d shows similar results using default model 2. Since this model is very close to the exact result, iterations now lead to a small improvement, but even with such an accurate default function there is an improvement. In Fig. 7, the curve “Iterated” shows results when iteration is allowed ( $N_{\text{iter}} \leq 40$ ). This leads to a substantial improvement in the accuracy. Fig. 7b shows similar results for default model 2.

#### V. QUADRATIC ENTROPY

In Eq. (6) in Sec. II we introduced the entropy, containing a logarithm. As a result, the basic equations of MaxEnt are nonlinear, which makes the analysis complicated. For this reason, we introduce a new definition of the entropy, which is used in this section for analyzing the behavior of MaxEnt. The expression in Eq. (5) is expanded to lowest (second) order in the deviation between the solution and the default function. We then *define* this as the entropy, and use it in *this* section. This is then not an approximation but simply a new method. This method has some problems. For instance, it is not guaranteed that the spectrum is positive. Therefore we do not recommend the use of this method for calculating spectra, but simply use it to analyze MaxEnt.

We define the entropy

$$S = -\frac{1}{2} \sum_{i=1}^{N_\omega} f_i \frac{(A_i^{(n)} - m_i^{(n)})^2}{m_i^{(n)}}, \quad (23)$$

where we have allowed for the possibility of the MaxEnt calculation being iterated, i.e.,  $m^{(n)}$  depends on the iteration  $n$ . The original default function is  $m^{(0)}$ . This leads to the equations

$$\begin{aligned} a &= K^T \sigma^{-2} \bar{G} + \alpha f \\ b^{(n)} &= K^T \sigma^{-2} K + \frac{\alpha f}{m^{(n)}} \end{aligned} \quad (24)$$

where matrix notations have been used and  $\alpha f$  and  $\alpha f/m^{(n)}$  are diagonal matrices. Then  $A^{(n+1)} = [(b^{(n)})^{-1} a]$ . The error in  $A^{(n+1)}$  is

$$\Delta A^{(n+1)} = [b^{(n)}]^{-1} [K^T \sigma^{-2} \Delta \bar{G} + \alpha f \frac{\Delta m^{(n)}}{m^{(n)}}], \quad (25)$$

where  $\Delta m^{(n)} = m^{(n)} - A^{\text{exact}}$ . We define

$$\begin{aligned} \delta G^{(n)} &= \sqrt{\frac{m^{(n)}}{\alpha f}} K^T \sigma^{-2} \Delta \bar{G} \\ \delta A^{(n)} &= \sqrt{\frac{\alpha f}{m^{(n)}}} \Delta A^{(n)} \\ c^{(n)} &= \sqrt{\frac{\alpha f}{m^{(n)}}} [b^{(n)}]^{-1} \sqrt{\frac{\alpha f}{m^{(n)}}} \end{aligned} \quad (26)$$

and a similar definition for  $\delta m^{(n)}$  as for  $\delta A^{(n)}$ .  $\sum_i [\delta A_i^{(n)}]^2$  contains the same integration factor  $f_i$  as has been used earlier, but due to the factor  $1/m^{(n)}$  it gives more weight to errors where  $m^{(n)}$  is small. We have

$$\delta A^{(n+1)} = c^{(n)}[\delta G^{(n)} + \delta m^{(n)}]. \quad (27)$$

We find the eigenvalues  $\varepsilon_\nu^{(n)}$  and eigenvectors  $|\nu^{(n)}\rangle$  of the symmetric matrix  $c^{(n)}$ . Introducing the expansion in these eigenvectors,  $\delta A_\nu^{(n)} = \langle \nu^{(n)} | \delta A^{(n)} \rangle$  we obtain

$$\delta A_\nu^{(n+1)} = \varepsilon_\nu^{(n)}[\delta G_\nu^{(n)} + \delta m_\nu^{(n)}]. \quad (28)$$

The matrix  $c$  can be rewritten as

$$c^{(n)} = \left[ \sqrt{\frac{m^{(n)}}{\alpha f}} K^T \sigma^{-2} K \sqrt{\frac{m^{(n)}}{\alpha f}} + 1 \right]^{-1} \quad (29)$$

For the cases we have considered, the first matrix inside the bracket has a broad range of positive eigenvalues, extending from eigenvalues much smaller than one to much larger than one. As a result, the matrix  $c^{(n)}$  is found to have some very small eigenvalues and many eigenvalues very close to one. This is illustrated in Table II, which shows the lowest eigenvalues for the default model 1 and  $\alpha = 40$ .

Fig. 8 shows the eigenfunctions  $|\nu^{(0)}\rangle$  corresponding to the lowest eigenvalues in Table I. The lowest function is nodeless, and the higher functions have an increasing number of nodes. Functions with the eigenvalue very close to one oscillate so rapidly that the corresponding components of  $\delta m_\nu^{(n)}$  and  $\delta G_\nu^{(n)}$  tend to have small weights, as shown in Table II. As a comparison, Table II also shows the expansion coefficients of the default model 2 in the eigenfunctions obtained for the default model 1 and  $\alpha = 40$ .

It is crucial for the success of a MaxEnt calculation that the coefficients  $\delta m_\nu^{(n)}$  and  $\delta G_\nu^{(n)}$  are typically small for  $\varepsilon_\nu^{(n)}$  close to one. From Eq. (28) it follows that errors  $\delta m_\nu^{(n)}$  and  $\delta G_\nu^{(n)}$  corresponding to eigenvalues  $\varepsilon_\nu$  much smaller than one give a strongly reduced contribution to the error  $\delta A_\nu^{(n+1)}$ , while errors corresponding to the eigenvalue one are not reduced at all. For these components the deviation of the default model from the true result are taken over completely.

At the same time this sets the limits for MaxEnt calculations. A MaxEnt calculation fails if  $A(\omega)$  has structures on such a small energy scale that there are important expansion coefficients  $\delta A_\nu^{(n+1)}$  corresponding to eigenvalues close to one, since the MaxEnt calculation gives no additional information about these components. This also shows the danger of putting in too much structure on a small energy scale in the default function. This would make components  $\delta m_\nu^{(n)}$  corresponding to  $\varepsilon_\nu \approx 1$  important and the MaxEnt calculation would not remove them from  $A_\nu^{(n+1)}$ , even if there is no support for such components in the data.

The results in Figs. 5 and 6 show a beating pattern, where the different calculations agree approximately for certain values of  $\omega$ . This must be related to the noise in the input data, since this is what differs between the calculations. The reason can be seen from Table II and Fig. 8. The contribution of the noise to the output is given by  $\varepsilon_\nu \delta G_\nu^{(n)}$ . This contribution comes mainly from the eighth and ninth eigenvalues. The corresponding eigenfunctions in Fig. 8 have their zeros approximately where the deviations between the calculations in Fig. 5 are small, although the agreement is not perfect. The reason is probably the nonlinearity due to the logarithm in Eq. (11). For instance, if the logarithm is expanded to second order, the resulting product of two functions generates functions with more nodes than either of the two functions. As a result we find that  $\delta A_\nu^{(n+1)}$  has appreciable errors also for components with a few more nodes than the eighth and ninth eigenfunctions. This then shifts the beating pattern slightly towards lower energies.

We introduce the projection operator

$$P^{(n)} = \sum_\nu |\nu^{(n)}\rangle \langle \nu^{(n)}| \Theta(\varepsilon_0 - \varepsilon_\nu^{(n)}), \quad (30)$$

where the  $\Theta$ -function selects states with eigenvalues smaller than  $\varepsilon_0 < 1$ . Eq. (28) can now be iterated. If we assume that  $\varepsilon_\nu^{(n)}$  is independent of  $n$ , which is a good approximation, we obtain

$$\delta A_\nu^{(n+1)} = \begin{cases} \sum_{i=1}^{n+1} \varepsilon_\nu^i \delta G_\nu^{(0)} + \varepsilon_\nu^{n+1} \delta m_\nu^{(0)} & \text{for } \varepsilon_\nu \leq \varepsilon_0 \\ \varepsilon_\nu [\delta G_\nu^{(0)} + \delta m_\nu^{(0)}] & \text{for } \varepsilon_\nu > \varepsilon_0 \end{cases} \quad (31)$$

This illustrates how iteration reduces the systematic error for components with  $\varepsilon_\nu \leq \varepsilon_0$ , but increases the statistical error. Whether iteration pays off then depends on the relative size of the statistical and systematic errors and the choice of  $\varepsilon_0$ . In this linearized version, however, it does not pay off to include all states in the projection operator (leading to  $P^{(n)} \equiv 1$ ).

For the nonlinear case, the behavior is a bit different. From the expression for the error in Eq. (11), it follows that  $\ln(m/A^{\text{exact}})$  enters. Expanding the logarithm leads to terms with products of eigenfunctions of the type in Fig. 8. Such products couple to higher eigenfunctions with more nodes. The result is that the error of a certain  $\nu$ -component of  $\ln(m/A^{\text{exact}})$  depends not only on the error of that  $\nu$ -component of  $m$  but also on the errors of other components, in particular lower ones. Whether the errors from the different contributions add constructively or destructively depends on the specifics of the model. For the cases we considered the contributions to the higher components often add destructively. Then it can be more favorable to iterate all components rather than just the ones that would be favorable according to Eq. (31). For the cases we have studied, this has usually been the case and this is the approach we used in Sec. IV.



TABLE II: Lowest eigenvalues  $\varepsilon_\nu$  of the matrix  $c^{(0)}$  [Eq. (26)] and the corresponding amplitudes  $\delta m_\nu^{(0)}$  (1) and  $\delta G_\nu^{(0)}$  for default model 1. The values of  $\delta m_\nu^{(0)}$  (2) for default model 2 (expanded in the functions corresponding to default model 1) and the expansion coefficients of  $\sigma(\omega)$  are also shown. The larger eigenvalues are all close to unity, the amplitudes of the corresponding  $\delta G_\nu^{(0)}$  are all smaller than  $2 \cdot 10^{-4}$ . The corresponding values of  $\delta m_\nu^{(0)}$  are also fairly small, smaller than 0.1 for default model 1 and smaller than 0.02 for default model 2. We used  $\beta = 15$ ,  $N_\omega = 121$ ,  $\omega_{\max} = 12$ ,  $\alpha = 40$  and  $\sigma = 0.001$ .

$\varepsilon_\nu$	$.3 \cdot 10^{-6}$	$.2 \cdot 10^{-5}$	$.6 \cdot 10^{-5}$	$.30 \cdot 10^{-4}$	$.2 \cdot 10^{-3}$	.002	.016	.166	0.747	0.982	0.999	1.000	1.000
$\delta m_\nu^{(0)}$ (1)	-.92	-.81	-.28	-1.7	1.1	-.01	-1.1	-.17	.31	-.08	-.33	-.15	.007
$\delta m_\nu^{(0)}$ (2)	.80	-.32	.06	-.08	.08	.03	-.19	-.11	.07	.03	-.10	-.08	-.007
$\sigma_\nu$	3.1	-2.4	2.9	1.5	-1.3	.09	1.0	.18	-.31	.08	.33	.16	-.007
$\sqrt{\langle (\delta G_\nu^{(0)})^2 \rangle}$	$1.6 \cdot 10^3$	675	371	176	65	23	7.90	2.1	.55	0.14	.03	.006	.001
$\varepsilon_\nu \sqrt{\langle (\delta G_\nu^{(0)})^2 \rangle}$	$.6 \cdot 10^{-3}$	.002	.002	.005	.014	.038	.128	.349	.411	.134	.030	.006	.001

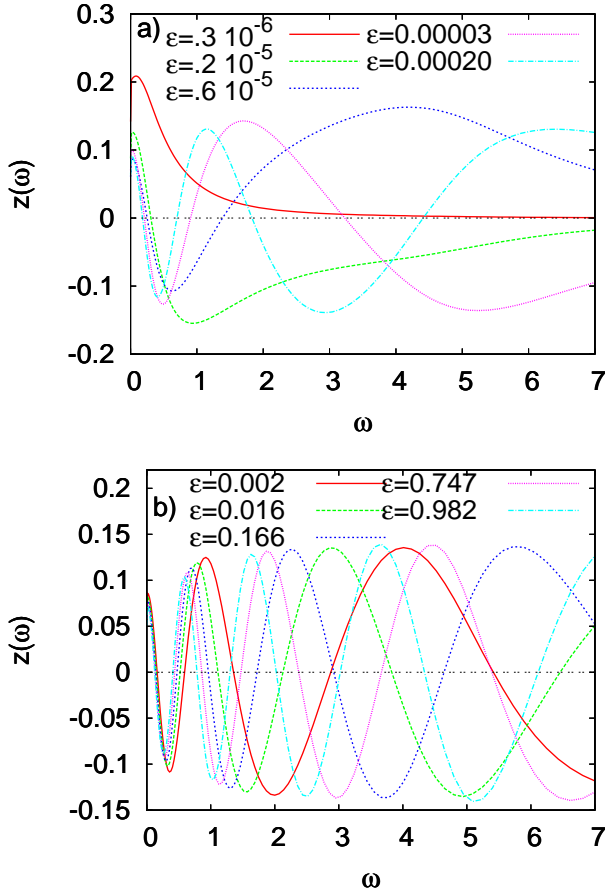


FIG. 8: (color on-line) Eigenfunctions to the matrix  $c^{(n)}$  in Eq. (26) corresponding to the 10 lowest eigenvalues. The figure illustrates how eigenfunctions corresponding to an eigenvalue close to one oscillate very rapidly.

## VI. SUMMARY

We have analyzed the MaxEnt approach for analytical continuation, defining a statistical error, due to noise in the imaginary axis input data, and a systematic error, due to errors in the default function entering the entropy.

The classical method for choosing the weight  $\alpha$  of the entropy can lead to a nonoptimal choice, reducing the systematic error at the cost of making the statistical error unnecessarily large. We find that the statistical error can be reduced by splitting up the data in batches. A MaxEnt calculation is performed for each batch and the result is averaged. This approach increases the systematic error but the total error can be reduced. We have also studied an iterative approach, where the output spectrum is used as default function in a new MaxEnt calculation. We find that a straightforward application of this approach often gives worse results due to a rapid increase of the statistical error. By splitting up the data in batches, the statistical error can be reduced sufficiently that this is less serious. The reduction of the systematic error can then outweigh the increase of the statistical error.

To analyze MaxEnt method, we have studied a linearized version of the problem. In this formalism it is easier to see how the statistical error propagates, in particular in the case of iterations. One can also see how certain deviations of the default function from the exact result have little influence on the output, while others fully show up in the output. This illustrates the danger of having a default function with too much structure.

While this paper shows the potential for improving the MaxEnt method, it is harder to provide prescriptions for how to use this. In Sec. III we provided a prescription for how to split the data in batches, which we have found to often work fairly well for a give value of  $\alpha$ . This method makes the resulting error less sensitive to the optimization of alpha. Alternatively, one can simply split the data in, say 10, batches. For each batch the classical method of determining  $\alpha$  is used and the resulting MaxEnt results are averaged. This approach typically improves the accuracy of the output spectrum. In particular, it reduces the risk of finding spurious structures due to overfitting of noisy data, while some real structures can be lost in this approach.

## VII. ACKNOWLEDGMENTS

We would like to thank M. Jarrell for making his Max-Ent program available. One of us (GS) wants to thank

for support through the FWF “Lise-Meitner” grant n. M1136.

- 
- <sup>1</sup> M. Jarrell and J.E. Gubernatis, Phys. Rep. **269**, 133 (1996).
  - <sup>2</sup> R. N. Silver, D. S. Sivia, and J. E. Gubernatis, Phys. Rev. B **41**, 2380 (1990); J. E. Gubernatis, M. Jarrell, R. N. Silver, and D. S. Sivia, Phys. Rev. **44**, 6011 (1991); W. von der Linden, Appl. Phys. A **60**, 155 (1995).
  - <sup>3</sup> H. J. Vidberg and J. W. Serene, J. Low Temp. Phys. **29**, 179 (1977).
  - <sup>4</sup> G. A. Baker, Jr., *Essentials of Padé approximants* Academic, New York, 1975, p 100ff.
  - <sup>5</sup> C.E. Creffield, E.G. Klepfish, E.R. Pike, and S. Sarkar, Phys. Rev. Lett. **75**, 517 (1995).
  - <sup>6</sup> I.S. Krivenko and A.N. Rubtsov, arXiv:cond-mat/0612233
  - <sup>7</sup> A. S. Mishchenko, N. V. Prokof'ev, A. Sakamoto, and B. V. Svistunov, Phys. Rev. B **62**, 6317 (2000).
  - <sup>8</sup> K. Vafayi and O. Gunnarsson, Phys. Rev. B **76**, 035115 (2007).